

Державний вищий навчальний заклад  
Донецький національний технічний університет  
Кафедра прикладної математики та інформатики

**«ЗАТВЕРДЖУЮ»**

Перший проректор

\_\_\_\_\_ Леонід БАЧУРІН

«\_\_\_\_\_» \_\_\_\_\_ 2023 р.

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**  
**ОК5 МЕТОДИ ТА ЗАСОБИ ОБРОБКИ ТЕКСТОВОЇ ІНФОРМАЦІЇ**  
(шифр і назва навчальної дисципліни)

Рівень освіти: другий (магістерський)

Спеціальність: 122 Комп'ютерні науки  
(шифр і назва спеціальності)

Освітня програма: Комп'ютерні науки  
(назва освітньої програми)

Мова навчання: українська

Робоча програма навчальної дисципліни «Методи та засоби обробки текстової інформації» для здобувачів вищої освіти за спеціальністю 122 Комп'ютерні науки «30» серпня 2023 р. – 8 с.

Розробник: Ярош І.В., старший викладач кафедри прикладної математики та інформатики.

Робоча програма затверджена на засіданні кафедри прикладної математики та інформатики

Протокол № 8 від «31» серпня 2023 р.

Зав. кафедри \_\_\_\_\_ ( Маслова Н.О. )

«31» серпня 2023 р.

Схвалено науково-методичною комісією з галузі знань 12 «Інформаційні технології»

Протокол № 5 від «01» вересня 2023 р.

Голова \_\_\_\_\_ ( Башков Є.О. )

## 1. Загальна інформація

Форма навчання	Денна	Заочна
Статус	Обов'язкова	
Обсяг в кредитах ЄКТС	7	7
Обсяг в годинах за навчальним планом, разом: в тому числі:	210	210
лекції:	32	8
практичні заняття:	32	8
лабораторні заняття:	–	–
семінари:	–	–
самостійна робота:	146	194
Форма підсумкового контролю	Екзамен	
Дисципліну викладають	Викладач Ярош Ірина Вікторівна, <a href="https://donntu.edu.ua/kitaer/pmi">https://donntu.edu.ua/kitaer/pmi</a> , <a href="mailto:iryna.yarosh@donntu.edu.ua">iryna.yarosh@donntu.edu.ua</a> , <a href="https://wiki.donntu.edu.ua/view/Ярош_Ірина_Вікторівна">https://wiki.donntu.edu.ua/view/Ярош_Ірина_Вікторівна</a>	

**Передумови для вивчення дисципліни:** теоретичною базою вивчення навчальної дисципліни є такі дисципліни: «Дискретна математика», «Теорія ймовірностей і математична статистика», «Чисельні методи», «Інтелектуальний аналіз даних».

## 2. Мета вивчення навчальної дисципліни «Методи та засоби обробки текстової інформації»

**Метою** вивчення навчальної дисципліни є формування системи теоретичних знань і практичних навичок з основ і принципів опрацювання інформації, поданої в текстовому вигляді.

Основні завданнями вивчення навчальної дисципліни – навчити студентів проводити синтаксичний, морфологічний із застосуванням затверджених підходів.

Як результат вивчення навчальної дисципліни повинні бути сформовані наступні **компетентності**:

- здатність вчитися і оволодівати сучасними знаннями;
- навички використання інформаційних і комунікаційних технологій;
- базові знання фундаментальних розділів математики в обсязі, необхідному для володіння математичним апаратом, потрібним для вивчення дисциплін технологічного спрямування;

- здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- здатність використовувати системні та базові знання зі штучного інтелекту, що включають в себе штучні нейронні мережі, розпізнавання образів та машинне навчання, експертні системи, нечітка логіка, агентні системи, розуміння природо-мовних текстів, автоматичне реферування;
- здатність застосовувати знання структури, функціонування, технічного та програмного забезпечення інтелектуальних та веб-орієнтованих комп'ютерних систем.

Як результат вивчення навчальної дисципліни повинні бути сформовані наступні **програмні результати навчання**:

- показати в галузі інформатики й сучасних інформаційних технологій, що включають в себе мови програмування, системи управління базами даних та знань, операційні системи, вебдизайн;
- продемонструвати вміння виконувати комп'ютерне моделювання процесів та явищ на стадії проєктування з використанням сучасних комп'ютерних систем;
- продемонструвати здатність генерувати нові ідеї і уміння обґрунтування нових інноваційних проєктів та просування їх на ринку;
- показати здатність зрозумілого і недвозначного донесення власних висновків, знань та пояснень до фахівців і нефахівців, зокрема і в процесі викладацької діяльності. Здатність зрозуміти роботу інших, давати і отримувати чіткі інструкції.;
- показати здатність до самостійного вирішення поставлених задач інноваційного характеру (кваліфікаційна робота, курсове проєктування), уміння аргументувати і захищати отримані результати та прийняті рішення, зокрема і публічно
- показати уміння обґрунтування та оцінювання інноваційних проєктів, знання методик просування їх на ринку, вміння виконувати економетричну та науковометричну оцінки.

### **3. Очікувані результати навчання**

Результати навчання, які базуються на програмних результатах навчання:

- вміння проводити синтаксичний аналіз за допомогою БС-граматики, ТП-граматики, мережевої граматики, штучної нейронної мережі Елмана та RAAM, метода N-грам;
- вміння проводити морфологічний аналіз за допомогою алгебри скінченних предикатів, системи «Сенс-Текст»;
- вміння проводити морфологічний аналіз за допомогою продукційних правил, семантичної мережі, фреймів, сценаріїв, числення предикатів, штучної нейронної мережі XRAAM.

Опанування дисципліною надає базис для подальшого вивчення наступних дисциплін спеціальності.

#### 4. Засоби діагностики результатів навчання

Поточний контроль здійснюється під час проведення практичних занять і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Форма проведення поточного контролю – усна бесіда за результатами виконання практичних робіт.

Підсумковий контроль проводиться з метою оцінювання результатів навчання та визначається підсумками результатів виконання й захисту практичних робіт по кожному зі змістовних модулів та оцінка розрахункової роботи.

Підсумковий семестровий контроль – екзамен.

#### 5. Критерії оцінювання результатів навчання

Критерії оцінювання мають формулювати порядок оцінювання під час поточного контролю (за результатами практичних, лабораторних, семінарських занять та виконання індивідуальних або групових завдань) та підсумкового контролю.

Пр1	Пр2	Пр3	Пр4	Поточний контроль	Екзамен	Максимальний бал
10	10	10	10	40	60	100
6	6	6	6	24		

Примітка: 1) Пр1-Пр4 – практичні роботи; 2) у числівнику максимальний бал – при своєчасному та правильному виконанні, у знаменнику – мінімальний (при правильному, але несвоечасному виконанні); 3) в оцінку поточного контролю з виконання практичних робіт включено контрольні та поточні опитування

Результати підсумкового контролю оцінюються за 100-бальною шкалою та чотирибальною («відмінно», «добре», «задовільно», «незадовільно»).

Відповідність між шкалами встановлюється наступним чином:

Оцінка	
За 100-бальною шкалою	Для екзамену
90-100	відмінно
74-89	добре
60-73	задовільно
0-59	незадовільно

#### 6. Програма навчальної дисципліни

##### 6.1. Основні теми дисципліни

**Тема 1.** Словотвір.

**Тема 2.** Словозміна.

**Тема 3.** Синтаксис.

**Тема 4.** Морфологічний аналіз.

**Тема 5.** Синтаксичний аналіз.

**Тема 6.** Семантичний аналіз.

### **6.2. Теми практичних (семінарських) занять**

№ з/п	Назва теми	Кількість годин	
		д.ф.н.	з.ф.н.
1	Створення контекстно-залежної граматики Тема 3. Синтаксис Тема 4. Морфологічний аналіз	8	2
2	Створення трансформаційної породжуючої граматики Тема 4. Морфологічний аналіз	8	2
3	Створення розширеної мережі переходів Тема 5. Синтаксичний аналіз	8	2
4	Обчислення ймовірності пропозиції на підставі біграм і триграм Тема 6. Семантичний аналіз	8	2
<b>Усього годин</b>		32	8

### **6.3. Теми лабораторних занять**

Робочим навчальним планом лабораторні заняття не передбачені.

### **6.4. Самостійна робота**

№ з/п	Назва теми	Кількість годин	
		д.ф.н.	з.ф.н.
1	2	3	4
1	Словотвір. Основні поняття словотвору. Частини слова. Способи словотворення. Морфонологічні явища в словотворенні.	20	27
2	Словозміна. Частини мови і морфологічні категорії. Морфонологічні явища в словозміні.	20	27
3	Синтаксис. Синтаксичні одиниці та синтаксичні зв'язку. Формальна організація пропозиції. Смислова організація пропозиції. Синтаксичні одиниці та синтаксичні зв'язку. Формальна організація пропозиції. Смислова організація пропозиції. Базові субстанціональні синтаксеми. Вторинні субстанціональні синтаксеми. Базові предикатні синтаксеми. Вторинні предикатні синтаксеми.	20	27

1	2	3	4
4	Морфологічний аналіз. Декларативний метод. Процедурний метод. Комбінований метод (поєднання декларативного і процедурного). Метод заснований на універсальних морфологічних моделях. Системи рівнянь алгебри скінченних предикатів. Система «Сенс-Текст».	28	37
5	Синтаксичний аналіз. Детерміністський підхід. Граматика безпосередньо складових. Трансформаційна граматика, що породжує. Мережева граматика. Нейромережевий підхід. Процедура навчання (алгоритм зворотного поширення). Стохастичний підхід.	28	38
6	Семантичний аналіз. Продукційні правила. Семантична мережу. Фрейм. Сценарій. Обчислення предикатів. Алгебра кінцевих предикатів. Нейромережевий підхід. Реляційні моделі. Об'єднання морфологічного, синтаксичного і семантичного аналізів.	30	38
<b>Усього годин</b>		<b>146</b>	<b>194</b>

#### **6.4 Індивідуальні та/або групові завдання**

Робочим навчальним планом індивідуальні (розрахункові) роботи не передбачені.

### **7. Література**

#### **7.1.Основна**

1. Вавіленкова А. І. Теоретичні основи аналізу електронних текстів : монографія. – К. : ТОВ «СІК ГРУП УКРАЇНА», 2016. – 192 с.
2. Інтелектуальна обробка текстів : навчальний посібник / В. Ю. Тарануха. – Київ, 2014. – 80 с.
3. Manika Lamba, Margam Madhusudhan. Text Mining for Information Professionals : An Uncharted Territory. – Springer, 2022. – 372 p.

#### **7.2.Допоміжна**

1. Прикладна лінгвістика : навчальний посібник / Н. А. Цимбал. – Умань : Візаві, 2019. – 106 с.
2. Alexandra George. Python Text Mining : Perform Text Processing, Word Embedding, Text Classification and Machine Translation. BPB Publications, 2022. – 320 p.
3. Jackson P., Moulinier I. Natural language processing for online applications : text retrieval, extraction, and categorization. – John Benjamins Pub, 2012. – 225 p.
4. Julia Silge, David Robinson. Text Mining with R : A Tidy Approach. – O'Reilly Media, 2017. – 191 p.

5. Gabe Ignatow, Rada F. Mihalcea. An Introduction to Text Mining: Research Design, Data Collection, and Analysis. – SAGE Publications, Inc, 2017. – 346 p.
6. ChengXiang Zhai, Sean Massung. Text Data Management and Analysis : A Practical Introduction to Information Retrieval and Text Mining. – ACM Books, 2016. – 532 p.
7. Charu C. Aggarwal Machine Learning for Text. – Springer, 2022. – 588 p.
8. Chengqing Zong, Rui Xia, Jiajun Zhang. Text Data Mining. – Kindle Edition, 2021. – 363 p.
9. Rafael E. Banchs. Text Mining with MATLAB. – Kindle Edition, Springer, 2021. – 492 p.
10. Goldberg Y. Neural Network Methods for Natural Language Processing / Y. Goldberg. – Morgan & Claypool Publishers, 2017. – 309 p.
11. Reese R. M. Natural Language Processing with Java / R. M. Reese. – Packt Publishing, 2015. – 262 p.

### **7.3 Методична**

1. Конспект лекцій з дисципліни «Сучасні методи обробки природномовних текстів» для студентів спеціальності 122 Комп'ютерні науки денної та заочної форми навчання / Є. Є. Федоров, І. В. Ярош, Т. О. Черняк. – Покровськ : ДВНЗ «ДонНТУ», 2018. – 60 с. – Режим доступу: <http://ea.donntu.edu.ua:8080/jspui/handle/123456789/29087>.

### **8. Інформаційні ресурси**

1. Getting started with NLP [Електронний ресурс]. – Режим доступу: <https://medium.com/@gon.esbuyo/get-started-with-nlp-part-ii-overview-of-an-nlp-workflow-7ba1f5948b24>.
2. Никоненко А. О. Огляд комп'ютерно-лінгвістичних методів обробки природномовних текстів [Електронний ресурс]. – Режим доступу: <http://dspace.nbu.gov.ua/bitstream/handle/123456789/59840/15-Nykonenko.pdf?sequence=1>.
3. Концептуальні алгоритми виокремлення морфем для реалізації інформаційної технології обробки природномовних текстів [Електронний ресурс] – Режим доступу: <http://www.hups.mil.gov.ua/periodic-app/article/8322>.